BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Statistical Inference-I

Course Code: 23ST09DEC

Unit-V

BAYESIAN ESTIMATION

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

Ms. N. Saranya
Guest Faculty
Department of Statistics

Unit-V

BAYESIAN ESTIMATION

Bayesian estimation is a statistical method that helps someone deal with conditional probability. It is done by using prior evidence to estimate an unknown population parameter.

Bayesian statistics is based on the idea that any unknown quantity, such as a parameter, a hypothesis, or a prediction, can be described by a probability distribution that reflects our degree of belief about it. This probability distribution is called the posterior distribution, and it is updated whenever we observe new data using Bayes' theorem.

Objective Bayesian statistics

Non-informative priors are the workhorse of objective Bayesian statistics.

In general, the prior reflects the statistician's **subjective beliefs**, as well as **knowledge** accumulated before observing the data. However, there are many cases in which not only we have little prior knowledge, but we would also like not to rely on subjective beliefs.

- We aim to publish our analyses in a scientific journal;
- We are presenting results to a regulator or to another public body.

Uninformative priors are used to make Bayesian inferences as objective as possible.

What is prior?

Prior probability is **the likelihood of an event occurring before we see the evidence (data)**. In Bayesian Inference, the prior is our first estimate of probability based on what we know now, before additional evidence (new data) becomes available.

The parameter as a random variable

So far we have seen the *frequentist* approach to statistical inference

i.e. inferential statements about θ are interpreted in terms of repeat sampling.

In contrast, the Bayesian approach treats θ *as a random variable* taking values in Θ.

The investigator's information and beliefs about the possible values for θ, before any observation of data, are summarised by a **prior distribution** $\pi(\theta)$.

When data $\mathbf{X}=\mathbf{x}$ are observed, the extra information about θ is combined with the prior to obtain the **posterior distribution** $\pi(\theta \mid \mathbf{x})$ for θ given $\mathbf{X}=\mathbf{x}$.

Prior and posterior distributions

- By Bayes' theorem,
$$\pi(\theta|\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x} \mid \theta)\pi(\theta)}{f_{\mathbf{X}}(\mathbf{x})},$$
where $f_{\mathbf{X}}(\mathbf{x}) = \int f_{\mathbf{X}}(\mathbf{x}|\theta)\pi(\theta)d\theta$ for continuous θ, and $f_{\mathbf{X}}(\mathbf{x}) = \sum f_{\mathbf{X}}(\mathbf{x}|\theta_i)\pi(\theta_i)$ in the discrete case.
- Thus
$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto f_{\mathbf{X}}(\mathbf{x}|\theta)\pi(\theta) \qquad (1) \\ \text{posterior} &\propto \text{likelihood} \times \text{prior}, \end{aligned}$$
where the constant of proportionality is chosen to make the total mass of the posterior distribution equal to one.
- In practice we use (1) and often we can recognise the family for $\pi(\theta \mid \mathbf{x})$.

Bayesian approach to point estimation

- Let $L(\theta, a)$ be the loss incurred in estimating the value of a parameter to be a when the true value is θ.
- Common loss functions are quadratic loss $L(\theta, a) = (\theta - a)^2$, absolute error loss $L(\theta, a) = |\theta - a|$, but we can have others.
- When our estimate is a, the expected posterior loss is $h(a) = \int L(\theta, a)\pi(\theta|\mathbf{x})d\theta$.
- The **Bayes estimator $\hat{\theta}$ minimises the expected posterior loss**.
- For **quadratic loss**
$$h(a) = \int (a - \theta)^2 \pi(\theta|\mathbf{x})d\theta.$$
- $h'(a) = 0$ if
$$a \int \pi(\theta|\mathbf{x})d\theta = \int \theta\pi(\theta|\mathbf{x})d\theta.$$
- So $\hat{\theta} = \int \theta\pi(\theta|\mathbf{x})d\theta$, the **posterior mean**, minimises $h(a)$.

- For **absolute error loss**,

$$\begin{aligned} h(a) &= \int |\theta - a|\pi(\theta|\mathbf{x})d\theta = \int_{-\infty}^{a} (a-\theta)\pi(\theta|\mathbf{x})d\theta + \int_{a}^{\infty} (\theta - a)\pi(\theta|\mathbf{x})d\theta \\ &= a\int_{-\infty}^{a} \pi(\theta|\mathbf{x})d\theta - \int_{-\infty}^{a} \theta\pi(\theta|\mathbf{x})d\theta \\ &\quad + \int_{a}^{\infty} \theta\pi(\theta|\mathbf{x})d\theta - a\int_{a}^{\infty} \pi(\theta|\mathbf{x})d\theta \end{aligned}$$

  Now $h'(a) = 0$ if

$$\int_{-\infty}^{a} \pi(\theta|\mathbf{x})d\theta = \int_{a}^{\infty} \pi(\theta|\mathbf{x})d\theta.$$

- This occurs when each side is 1/2 (since the two integrals must sum to 1) so $\hat{\theta}$ is the **posterior median**.

1 Bayes Formula

For events A and B, recall that the conditional probability is

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A),$$

or

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Now, if we set

$$A = \{\theta = \theta_0\} \quad \text{and} \quad B = \{X = \mathbf{x}\},$$

then

$$P\{\theta = \theta_0|X = \mathbf{x}\} = \frac{P\{X = \mathbf{x}|\theta = \theta_0\}P\{\theta = \theta_0\}}{P\{X = \mathbf{x}\}}.$$

If the appropriate densities exist, then we can write **Bayes formula** as

$$f_{\Theta|X}(\theta_0|\mathbf{x}) = \left(\frac{f_{X|\Theta}(\mathbf{x}|\theta_0)}{\int f_{X|\Theta}(\mathbf{x}|\tilde{\theta})\pi(\tilde{\theta})\, d\tilde{\theta}}\right)\pi(\theta_0),$$

to compute the **posterior density** $f_{\Theta|X}(\theta_0|x)$ as the product of the **Bayes factor** and the **prior density**.

If T is a sufficient statistic and $f_{X|\Theta}(\mathbf{x}|\tilde{\theta}) = h(\mathbf{x})g(\theta, T(\mathbf{x}))$, then the Bayes factor

$$\frac{f_{X|\Theta}(\mathbf{x}|\theta_0)}{\int f_{X|\Theta}(\mathbf{x}|\tilde{\theta})f_{\Theta}(\tilde{\theta})\, d\tilde{\theta}} = \frac{h(\mathbf{x})g(\theta, T(\mathbf{x}))}{\int h(\mathbf{x})g(\tilde{\theta}, T(\mathbf{x}))\, d\tilde{\theta}} = \frac{g(\theta, T(\mathbf{x}))}{\int g(\tilde{\theta}, T(\mathbf{x}))\, d\tilde{\theta}}$$

is a function of T.

2 Bayes Action

Recall that given a loss function $\mathcal{L}$ and an estimator d the risk function $\mathcal{R} : \Theta \times \mathcal{D} \to \mathbb{R}$ is the expected loss for that decision.

$$\mathcal{R}(\theta, d) = E_\theta \mathcal{L}(\theta, d(X))$$

and the mean risk, or **Bayes risk**,

$$r(\pi, d) = \int_\Theta \mathcal{R}(\theta, d)\pi(\theta)\ d\theta = \int_\Theta \int_{\mathbb{R}^n} \mathcal{L}(\theta, d(x)) f_X(\mathbf{x}|\theta)\pi(\theta)\ d\mathbf{x}d\theta.$$

The decision function that minimizes risk is called the **Bayes action**.

If the loss function is $\mathcal{L}_1(\theta, a) = |\theta - a|$, then the posterior median minimizes risk and thus the Bayes action $\hat{\theta}_1(\mathbf{x})$ satisfies

$$\frac{1}{2} = \int_{-\infty}^{\hat{\theta}_1(\mathbf{x})} f_{\Theta|X}(\theta|x)\ d\theta.$$

If the loss function is $\mathcal{L}_2(\theta, a) = (\theta - a)^2$, then the posterior mean minimizes risk and thus the Bayes action

$$\hat{\theta}_2(\mathbf{x}) = E[\theta|X = \mathbf{x}] = \int \theta f_{\Theta|X}(\theta|\mathbf{x})\ d\theta.$$

For the example of a normal prior and normal observations, $\hat{\theta}_1(\mathbf{x}) = \hat{\theta}_2(\mathbf{x}) = \tilde{\theta}(\mathbf{x})$.

Definition: Informative and non-informative prior distribution

Definition: Let $p(\theta|m)$ be a prior distribution for the parameter θ of a generative model m. Then,

- the distribution is called an "informative prior", if it biases the parameter towards particular values;
- the distribution is called a "weakly informative prior", if it mildly influences the posterior distribution;
- the distribution is called a "non-informative prior", if it does not influence the posterior hyperparameters.

Types of uninformative priors

We will briefly describe below the following classes of non-informative priors:

- Bayes-Laplace uniform prior;
- Jeffreys' prior;
- Jaynes' maximum entropy prior
- Bernardo's reference prior.

What is an informative prior?

An informative prior is a probability distribution that reflects your existing knowledge or beliefs about a parameter before observing any data. For example, if you are estimating the proportion of voters who support a certain candidate, you might have some prior information from polls, surveys, or historical trends that can help you shape your prior. An informative prior can be contrasted with a non-informative prior, which is a distribution that expresses no or minimal information about the parameter, such as a uniform or vague distribution.

What is a normal conjugate prior?

If you have a conjugate prior this means that the prior comes from the same family of distributions and there is a closed-form solution for such problem, so the posterior distribution is directly available. This is exactly the case when you use normal prior for mean parameter of normal distribution.

Definition of a conjugate prior

We can now define the concept of a conjugate prior.

Definition Let Φ be a parametric family. A prior $p(\theta)$ belonging to Φ is said to be conjugate for the likelihood $p(x|\theta)$ if and only if the posterior $p(\theta|x)$ belongs to Φ.

- In Bayesian inference, the prior distribution of a parameter and the likelihood of the observed data are combined to obtain the posterior distribution of the parameter.
- If the prior and the posterior belong to the same parametric family, then the prior is said to be conjugate for the likelihood.

Prior, likelihood and posterior

In a Bayesian inference problem, we specify two distributions:

- the likelihood, that is, the distribution of the observed data x conditional on the parameter θ:

$$p(x|\theta)$$

- the prior distribution of the parameter

$$p(\theta)$$

After observing the data, we use Bayes' rule to compute the posterior distribution

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int_\theta p(x|\theta)p(\theta)d\theta}$$

Bernoulli likelihood and beta priors

Remember that a Bernoulli random variable is equal to 1 with probability q and to 0 with probability $1-q$.

Suppose that we observe a realization x of the Bernoulli variable and we want to carry out some Bayesian inference on the unknown parameter q.

The likelihood has exponential form:

$$\begin{aligned}
& p(x|q) \\
&= 1_{\{x\in\{0,1\}\}} q^x (1-q)^{1-x} \\
&= 1_{\{x\in\{0,1\}\}} \left(\frac{q}{1-q}\right)^x (1-q) \\
&= 1_{\{x\in\{0,1\}\}} \exp\left[x \ln\left(\frac{q}{1-q}\right) + \ln(1-q)\right] \\
&= 1_{\{x\in\{0,1\}\}} \exp\left[x \ln\left(\frac{q}{1-q}\right) - \ln\left(1+\exp\left(\ln\left(\frac{q}{1-q}\right)\right)\right)\right] \\
&= h(x)\exp[\theta^\top T(x) - A(\theta)] \\
&= p(x|\theta)
\end{aligned}$$

where $1_{\{x\in\{0,1\}\}}$ is an indicator function equal to 1 if $x \in \{0,1\}$ and to 0 otherwise, and

$$\begin{aligned}
\theta &= \ln\left(\frac{q}{1-q}\right) \\
T(x) &= x \\
A(\theta) &= \ln[1+\exp(\theta)] \\
h(x) &= 1_{\{x\in\{0,1\}\}}
\end{aligned}$$

The natural family of conjugate priors contains priors of the form

$$\begin{aligned}
p(\theta) &\propto \exp[\theta\chi - \nu A(\theta)] \\
&= \exp[\theta\chi - \nu \ln[1+\exp(\theta)]]
\end{aligned}$$

Since θ is an increasing function of q and

$$\begin{aligned}
\frac{d\theta}{dq} &= \frac{d}{dq}\ln\left(\frac{q}{1-q}\right) \\
&= \frac{1-q}{q}\left[\frac{1}{1-q} + \frac{q}{(1-q)^2}\right] \\
&= \frac{1-q+q}{q(1-q)} = \frac{1}{q(1-q)}
\end{aligned}$$

we can apply the formula for the density of an increasing function:

$$\begin{aligned}
p(q) &\propto \exp\left[\ln\left(\frac{q}{1-q}\right)\chi + \nu\ln(1-q)\right]\frac{1}{q(1-q)} \\
&= \left(\frac{q}{1-q}\right)^\chi (1-q)^\nu \frac{1}{q(1-q)} \\
&= q^{\chi-1}(1-q)^{\nu-\chi-1}
\end{aligned}$$

Thus, the natural family of conjugate priors contains priors that assign to q a Beta distribution with parameters χ and $\nu-\chi$.

According to the general formula derived above for natural families, the posterior distribution of θ is

$$\begin{aligned}
p(\theta|x) &\propto \exp[\theta^\top(\chi + T(X)) - (\nu+1)A(\theta)] \\
&= \exp[\theta(\chi+x) - (\nu+1)A(\theta)]
\end{aligned}$$

which implies (by the same argument just used for the prior) that the posterior distribution of q is

$$p(q|x) \propto q^{\chi+x-1}(1-q)^{\nu+1-\chi-x-1}$$

that is, a Beta distribution with parameters $\chi+x$ and $\nu-\chi+1-x$.

Bayes factor for testing Hypothesis

Bayes factors represent an informative alternative to P-values for reporting outcomes of hypothesis tests. They provide direct measures of the relative support that data provide to competing hypotheses and are able to quantify support for true null hypotheses

Bayesian hypothesis testing is then often done via the Bayes factor BF_{10}, the predictive updating factor which measures the change in relative beliefs about hypothesis H_1 relative to hypothesis H_0 given the data x:

$$\underbrace{\frac{p\left(x|H_1\right)}{p\left(x|H_0\right)}}_{BF_{10}(x)} = \underbrace{\frac{\mathbb{P}\left(H_1|x\right)}{\mathbb{P}\left(H_0|x\right)}}_{\text{Posterior odds}} \cdot \underbrace{\frac{\mathbb{P}\left(H_0\right)}{\mathbb{P}\left(H_1\right)}}_{\text{Prior odds}}$$

The Bayes factor BF_{10} therefore quantifies the evidence by indicating how much more likely the observed data are under the rival models. Note that the Bayes factor critically depends on the prior distributions assigned to the parameters in each of the models, as the parameter values determine the models' predictions.

values determine the models' predictions. It can also be rewritten as the ratio of posterior and prior odds. Bayesian parameter estimation for an unknown parameter θ in general is achieved by considering the posterior distribution $p(\theta|x)$ of the parameter after observing the data x:

$$p\left(\theta|x\right) = \frac{p\left(x|\theta\right) \cdot p\left(\theta\right)}{p(x)} \tag{2}$$

where $p(x|\theta)$ is the likelihood function, $p(\theta)$ the prior, and in most realistic settings, the marginal likelihood $p(x)$ in the denumerator cannot be calculated in closed form or is prohibitively effortful to compute. Therefore, Markov-Chain-Monte-Carlo (MCMC) algorithms have been developed in the last decades, alleviating the requirement of computing $p(x)$ from practitioners, because most MCMC algorithms only need a function *proportional* to the posterior to work, so that

$$p(\theta|x) \propto p(x|\theta) \cdot p(\theta) \tag{3}$$

Bayes factors represent the ratio of the marginal probability assigned to data by competing hypotheses and, when combined with prior odds assigned between hypotheses, yield an estimate of the posterior odds that each hypothesis is true. That is,

$$\textbf{posterior odds} = \textbf{ Bayes factor} \times \textbf{ prior odds},$$

or, more precisely,

$$\frac{P(H_1|\mathbf{x})}{P(H_0|\mathbf{x})} = \frac{m_1(\boldsymbol{x})}{m_0(\boldsymbol{x})} \times \frac{P(H_1)}{P(H_0)}.$$

Here, $P(H_i|\mathbf{x})$ denotes the posterior probability of hypothesis H_i given data $\mathbf{x}$; $P(H_i)$ denotes the prior probability assigned to H_i, and $m_i(\mathbf{x})$ denotes the marginal probability (or probability density function) assigned to the data under hypothesis H_i, for $i = 0$ (null) or $i = 1$ (alternative).

The marginal density of the data under the alternative hypothesis is given by

$$m_1(\mathbf{x}) = \int_\Theta f(\mathbf{x}|\theta)\pi_1(\theta)d\theta,$$

where $f(\mathbf{x}|\theta)$ denotes the sampling density of the data given an unknown parameter θ. In null hypothesis significance tests (NHSTs), the marginal density of the data under the null hypothesis, $m_0(\mathbf{x})$, is simply the sampling density of the data assumed under the null hypothesis. That is, if $H_0 : \theta = \theta_0$, then $m_0(\mathbf{x}) = f(\mathbf{x}|\theta_0)$. The function $\pi_1(\theta)$ represents the prior density for the parameter of interest θ under the alternative hypothesis, i.e., the alternative prior density.

Equivariance

We work in a local coordinate system. The second jet of L has the following coordinates induced via the local coordinates we are using:

1) the function L;
2) the first partial derivatives of L,

$$\frac{\partial L}{\partial q^i}, \qquad \frac{\partial L}{\partial v^i};$$

3) the second partial derivatives of L

$$\frac{\partial^2 L}{\partial q^i \partial q^j}, \qquad \frac{\partial^2 L}{\partial q^i \partial v^j}, \qquad \frac{\partial^2 L}{\partial v^i \partial v^j}.$$

Also $X = (v, v, a) \in E(TM)$ has components:

The Bayes factor is the ratio of two marginal likelihoods; that is, the likelihoods of two statistical models integrated over the prior probabilities of their parameters.[9]

The posterior probability $\Pr(M|D)$ of a model M given data D is given by Bayes' theorem:

$$\Pr(M|D) = \frac{\Pr(D|M)\Pr(M)}{\Pr(D)}.$$

The key data-dependent term $\Pr(D|M)$ represents the probability that some data are produced under the assumption of the model M; evaluating it correctly is the key to Bayesian model comparison.

5.3 The Pitman location estimator

The UMREE under squared error loss is

$$\delta(\boldsymbol{x}) = \delta_0(\boldsymbol{x}) - \mathrm{E}_0[\delta_0(\boldsymbol{X})|\boldsymbol{y}],$$

where $\delta_0(\boldsymbol{x})$ is an arbitrary equivariant estimator. The choice of δ_0 will not affect δ, it will only affect our ability to calculate $\mathrm{E}_0[\delta_0(\boldsymbol{X})|\boldsymbol{y}]$. Let's try a very simple equivariant estimator:

$$\delta_0(\boldsymbol{x}) = x_n.$$

Note that

$$\begin{aligned}\delta_0(g\boldsymbol{x}) &= \delta(\boldsymbol{x} + a) \\ &= x_n + a \\ &= \delta_0(\boldsymbol{x}) + a = \tilde{g}\delta_0(\boldsymbol{x}),\end{aligned}$$

and so this estimator is equivariant. Now we just need to calculate the conditional expectation:

$$\mathrm{E}[X_n|X_1 - X_n, \ldots, X_{n-1} - X_n].$$

The joint density of $(X_1 - X_n, \ldots, X_{n-1} - X_n, X_n) = (Y_1, \ldots, Y_{n-1}, X_n)$ can be found from the change of variables formula:

$$p_{y,x_n}(y_1, \ldots, y_{n-1}, x_n) = p_{\boldsymbol{x}}(x_1(\boldsymbol{y}, x_n), \ldots, x_{n-1}(\boldsymbol{y}, x_n), x_n)|d(\boldsymbol{y}, x_n)/d\boldsymbol{x}|$$

Now

- $|d(\boldsymbol{y}, x_n)/d\boldsymbol{x}| = 1$
- $x_i = y_i + x_n, \ i = 1, \ldots n-1.$

so we have

$$p_{y,x_n}(y_1, \ldots, y_{n-1}, x_n) = p_0(y_1 + x_n, \ldots, y_{n-1} + x_n, x_n).$$

Now recall $y_i = x_i - x_n$ for $i = 1, \ldots, n-1$:

$$\begin{aligned}\mathrm{E}[X_n|\boldsymbol{y}] &= \frac{\int (x_n - \theta) p_0(x_1 - \theta, \ldots, x_n - \theta)\ d\theta}{\int p_0(x_1 - \theta, \ldots, x_n - \theta)\ d\theta} \\ &= x_n - \frac{\int (x_n - \theta) p_0(x_1 - \theta, \ldots, x_n - \theta)\ d\theta}{\int p_0(x_1 - \theta, \ldots, x_n - \theta)\ d\theta} \\ \delta(\boldsymbol{x}) &= x_n - \mathrm{E}[X_n|\boldsymbol{y}] \\ &= \frac{\int \theta p_0(x_1 - \theta, \ldots, x_n - \theta)\ d\theta}{\int p_0(x_1 - \theta, \ldots, x_n - \theta)\ d\theta}.\end{aligned}$$

This is known as the Pitman estimator of θ. Note that it is equal to the generalized Bayes estimate under the prior measure $\pi(\theta) \propto 1, \theta \in \mathbb{R}$. Under this "prior,"

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &= \frac{\pi(\theta)p_0(x_1-\theta,\ldots,x_n-\theta)}{\int \pi(\theta)p_0(x_1-\theta,\ldots,x_n-\theta)\,d\theta} \\ &= \frac{p_0(x_1-\theta,\ldots,x_n-\theta)}{\int p_0(x_1-\theta,\ldots,x_n-\theta)\,d\theta}, \text{ so} \\ \mathrm{E}_\pi[\theta|\boldsymbol{x}] &= \frac{\int \theta p_0(x_1-\theta,\ldots,x_n-\theta)\,d\theta}{\int p_0(x_1-\theta,\ldots,x_n-\theta)\,d\theta}.\end{aligned}$$

4.1. Pitman method

Let $t_1,t_2,\ldots,t_n$ be a random sample of n observations from a population whose p.d.f is $f(t;\lambda)$; where $\lambda > 0$ is a scale parameter and $t_i > 0$. If $\hat{\lambda} = h(t_1,t_2,\ldots,t_n)$ is the estimator of the scale parameter λ, then $\hat{\lambda}$ should be as follows:

$$\hat{\lambda} = \frac{\int_0^\infty \frac{1}{\lambda^2} L(t_1,t_2,\ldots,t_n;\lambda)\,d\lambda}{\int_0^\infty \frac{1}{\lambda^3} L(t_1,t_2,\ldots,t_n;\lambda)\,d\lambda} \qquad (5)$$

The Pitman estimator $\hat{\lambda}$ is also a function of sufficient statistic. If the estimator $\hat{\lambda} = h(t_1,t_2,\ldots,t_n)$ is expressed as: $\hat{\lambda}_1 = h_1(ct_1, ct_2,\ldots ct_n) = c\,h_1(t_1,t_2,\ldots,t_n) = c\,h\,(t_1,t_2,\ldots,t_n) = c\hat{\lambda}$ .Then $\hat{\lambda}$ is scale invariant [17].

Now based on the above method, the likelihood function of exponential distribution is:

$$L(t_1,t_2,\ldots,t_n;\lambda) = \frac{1}{\lambda^n} e^{-\frac{\sum_{i=1}^n t_i}{\lambda}}$$

Therefore, the Pitman estimator $\hat{\lambda}_{Pitman}$ of the scale parameter λ will be as follows:

$$\hat{\lambda}_{Pitman} = \frac{\int_0^\infty \frac{1}{\lambda^{n+2}} e^{-\frac{\sum_{i=1}^n t_i}{\lambda}} d\lambda}{\int_0^\infty \frac{1}{\lambda^{n+3}} e^{-\frac{\sum_{i=1}^n t_i}{\lambda}} d\lambda}$$

Let $y = \frac{\sum_{i=1}^n t_i}{\lambda}$, $\Rightarrow \lambda = \frac{\sum_{i=1}^n t_i}{y}$, $\Rightarrow d\lambda = -\frac{\sum_{i=1}^n t_i}{y^2} dy$, then we get:

$$\hat{\lambda}_{Pitman} = \frac{\frac{1}{(\sum_{i=1}^n t_i)^{n+1}} \int_0^\infty y^n e^{-y}\,dy}{\frac{1}{(\sum_{i=1}^n t_i)^{n+2}} \int_0^\infty y^{n+1} e^{-y}\,dy}$$

$$\hat{\lambda}_{Pitman} = \frac{\frac{\Gamma(n+1)}{(\sum_{i=1}^n t_i)^{n+1}}}{\frac{\Gamma(n+2)}{(\sum_{i=1}^n t_i)^{n+2}}}$$

After simplification, we get a Pitman estimator for the parameter λ as follows:

$$\hat{\lambda}_{Pitman} = \frac{\sum_{i=1}^n t_i}{n+1} \qquad (6)$$

Therefore, using this estimator, the Pitman Reliability function of the exponential distribution will be given by:

$$\hat{R}(t)_{Pitman} = e^{-\frac{t}{\hat{\lambda}_{Pitman}}} \qquad (7)$$

One-tailed Tests

- A one-tailed test may be either left-tailed or right-tailed.
- A *left-tailed* test is used when the alternative hypothesis states that the true value of the parameter specified in the null hypothesis is less than the null hypothesis claims.
- A *right-tailed* test is used when the alternative hypothesis states that the true value of the parameter specified in the null hypothesis is greater than the null hypothesis claims

Two-tailed Tests

The main difference between one-tailed and two-tailed tests is that one-tailed tests will only have one critical region whereas two-tailed tests will have two critical regions. If we require a $(1-\alpha)$ 100% confidence interval we have to make some adjustments when using a two-tailed test.

Advantages of One-Tailed Tests

- They allow researchers to focus on a specific direction of the effect, which can be useful when there is a clear expectation about the direction.
- One-tailed tests have the potential to provide higher statistical power when the effect is expected to occur in the specified direction.
- They can be more efficient when resources are limited, as they concentrate the analysis on a specific direction.

Limitations of One-Tailed Tests

- One-tailed tests are sensitive only to changes occurring in the specified direction and may miss important effects in the opposite direction.
- They introduce a potential bias if the specified direction is chosen after observing the data, rather than being based on a priori expectations.
- Using a one-tailed test when there is no specific expectation about the direction can lead to misleading interpretations.

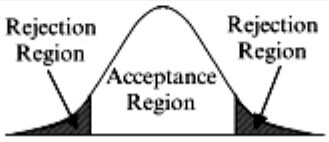
Advantages of Two-Tailed Tests

- Two-tailed tests allow researchers to detect any type of difference, regardless of the direction, providing a more comprehensive analysis.
- They are suitable when there is no specific expectation about the direction of the effect or when researchers want to avoid potential biases.
- Two-tailed tests provide a more conservative approach, considering both positive and negative differences.

Limitations of Two-Tailed Tests

- Two-tailed tests may have lower statistical power compared to one-tailed tests, as they distribute the power between both directions.
- They may require larger sample sizes to achieve the same level of statistical power as a one-tailed test.
- Two-tailed tests may provide less precise estimates of the effect size compared to one-tailed tests.

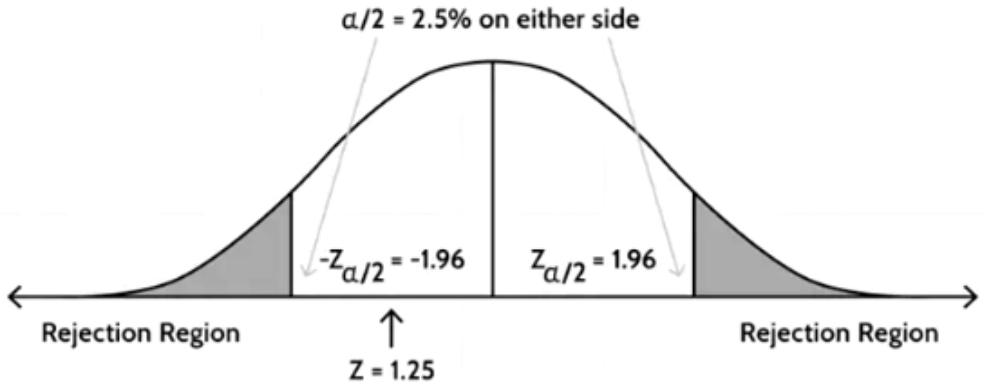
Understanding the advantages and limitations of each type of test allows researchers to make informed decisions and select the most appropriate approach for their specific research context.

One-Tailed Test (Left Tail)	Two-Tailed Test	One-Tailed Test (Right Tail)
$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X < \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X \neq \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X > \mu_0$
		

A sample of 100 clients of ABC is taken, and brokerage charges are calculated with the new rates of XYZ broker. If the mean of the sample is $18.75 and the sample standard deviation is $6, can any inference be made about the difference in the average brokerage bill between ABC and XYZ broker?

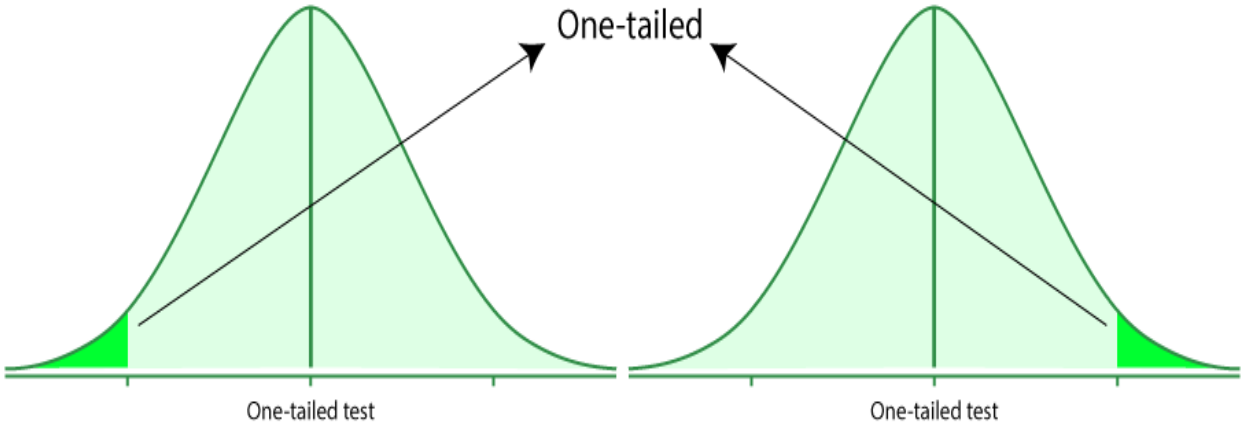
- H_0: Null Hypothesis: mean = 18
- H_1: Alternative Hypothesis: mean <> 18 (This is what we want to prove.)
- Rejection region: Z <= - $Z_{2.5}$ and Z>=$Z_{2.5}$ (assuming 5% significance level, split 2.5 each on either side).
- Z = (sample mean – mean) / (std-dev / sqrt (no. of samples)) = (18.75 – 18) / (6/(sqrt(100)) = 1.25

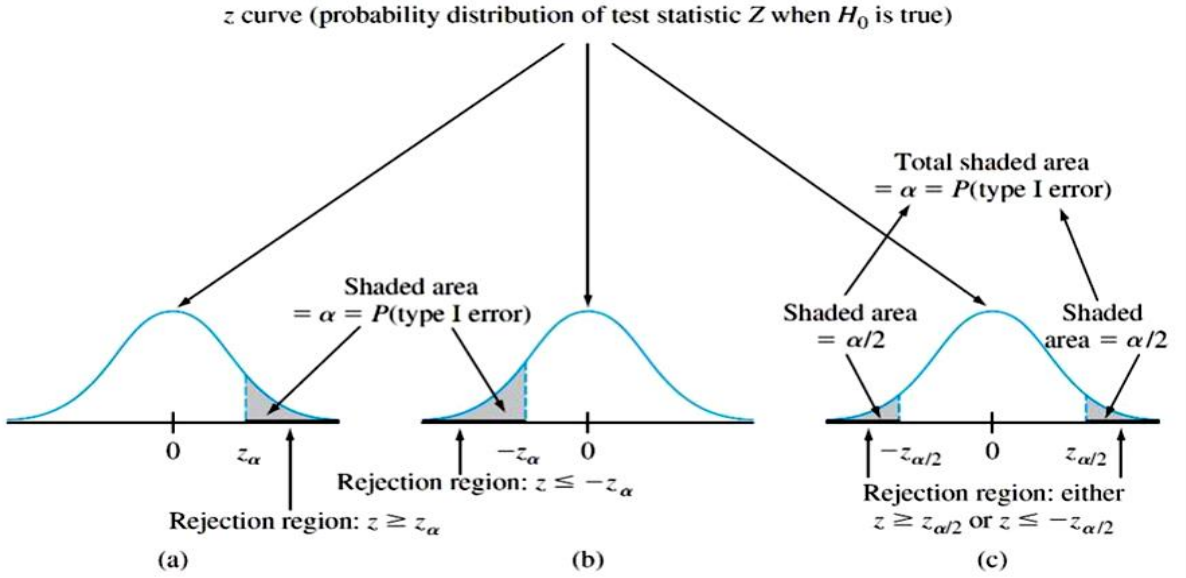
This calculated Z value falls between the two limits defined by: - $Z_{2.5}$ = -1.96 and $Z_{2.5}$ = 1.96.



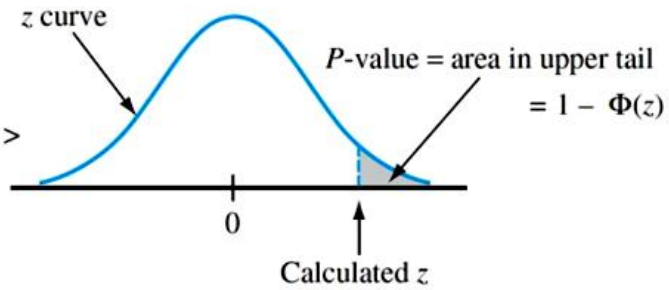
- **One and Two-Tailed Tests** are ways to identify the relationship between the statistical variables. For checking the relationship between variables in a **single direction** (Left or Right direction), we use a one-tailed test. A two-tailed test is used to check whether the relations between variables are in any direction or not.

- **One-Tailed Test:** A one-tailed test is based on a unit -directional hypothesis where the area of rejection is on only one side of the sampling distribution. It determines whether a particular population parameter is larger or smaller than the predefined parameter. It uses one single critical value to test the data.

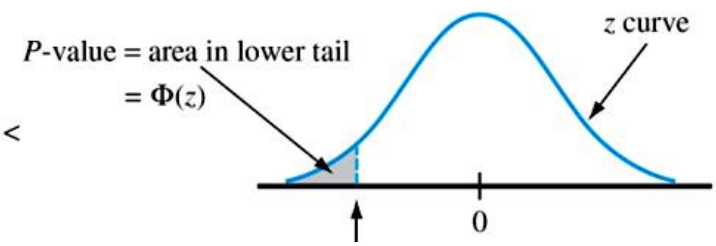




1. **Upper-tailed test**
   H_a contains the inequality >

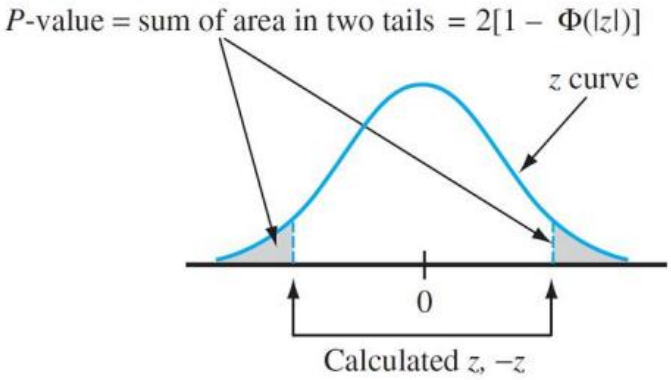


2. **Lower-tailed test**
   H_a contains the inequality <



P-Value for Z test

3. **Two-tailed test**
   H_a contains the inequality $\neq$



Difference Between One and Two-Tailed Test

One-Tailed Test	Two-Tailed Test
A test of any statistical hypothesis, where the alternative hypothesis is **one-tailed** either right-tailed or left-tailed.	A test of a statistical hypothesis, where the alternative hypothesis is **two-tailed**.
For one-tailed, we use either > or < sign for the alternative hypothesis.	For two-tailed, we use ≠ sign for the alternative hypothesis.
When the alternative hypothesis specifies a direction then we use a one-tailed test.	If no direction is given then we will use a two-tailed test.
Critical region lies entirely on either the right side or left side of the sampling distribution.	Critical region is given by the portion of the area lying in both the tails of the probability curve of the test statistic.
Here, the Entire **level of significance (α)** i.e. 5% has either in the left tail or right tail.	It splits the **level of significance (α)** into half.
Rejection region is either from the left side or right side of the sampling distribution.	Rejection region is from both sides i.e. left and right of the sampling distribution.
It checks the relation between the variable in a singles direction.	It checks the relation between the variables in any direction.
It is used to check whether the one mean is different from another mean or not.	It is used to check whether the two mean different from one another or not.

What is a Z-score?

A Z-score numerically describes a value's relationship to the mean of a group of values and is measured in terms of the number of standard deviations from the mean. If a Z-score is 0, it indicates that the data point's score is identical to the mean score whereas Z-scores of 1.0 and -1.0 would indicate values one standard deviation above or below the mean. In most large data sets, 99% of values have a Z-score between -3 and 3, meaning they lie within three standard deviations above and below the mean.

How to interpret z-Score?

Here is how to interpret z-scores:

- A z-score of less than 0 represents an element less than the mean.
- A z-score greater than 0 represents an element greater than the mean.
- A z-score equal to 0 represents an element equal to the mean.
- A z-score equal to 1 represents an element, which is 1 standard deviation greater than the mean; a z-score equal to 2 signifies 2 standard deviations greater than the mean; etc.
- A z-score equal to -1 represents an element, which is 1 standard deviation less than the mean; a z-score equal to -2 signifies 2 standard deviations less than the mean; etc.
- If the number of elements in the set is large, about 68% of the elements have a z-score between -1 and 1; about 95% have a z-score between -2 and 2 and about 99% have a z-score between -3 and 3.

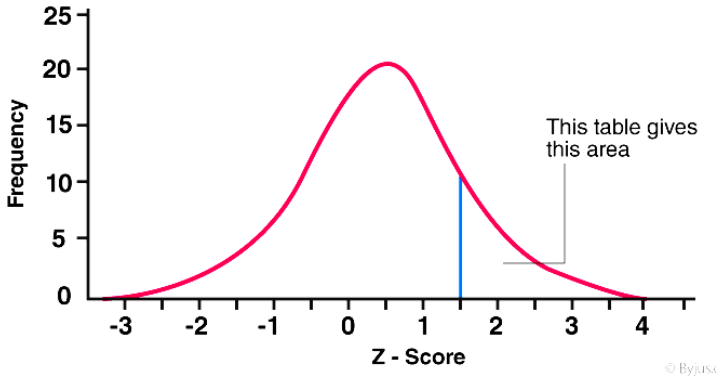
Z-Score Formula

- It is a way to compare the results from a test to a “normal” population.
- If X is a random variable from a normal distribution with mean (μ) and standard deviation (σ), its Z-score may be calculated by subtracting mean from X and dividing the whole by standard deviation.

$$Z = \frac{(X - \mu)}{\sigma}$$

where, x = test value
μ is mean and
σ is SD (Standard Deviation)

For the average of a sample from a population ‘n’, the mean is μ and the standard deviation is σ.



Example of Z score

Let us understand the concept with the help of a solved example:

Example: The test scores of students in a class test has a mean of 70 and with a standard deviation of 12. What is the probable percentage of students scored more than 85?

Solution: The z score for the given data is,

$$Z = (85-70)/12=1.25$$

From the z score table, the fraction of the data within this score is 0.8944. This means 89.44 % of the students are within the test scores of 85 and hence the percentage of students who are above the test scores of 85 = (100-89.44)% = 10.56 %.